---
title: "week_5"
author: "krishna sai surendra babu kalluri"
date: "11/07/2020"
output: html_document

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

1) The equation for the basis function is

$$ y_i =\beta_0 + \beta_1b_1(x_i)+\beta_2b_2(x_i)+\beta_3b_3(x_i)+...
\beta_k b_k(x_i)+E_i $$

The Basis function for polynomial function is

$$b_j(x_i)=x^j_i$$
The Basis function for step function is

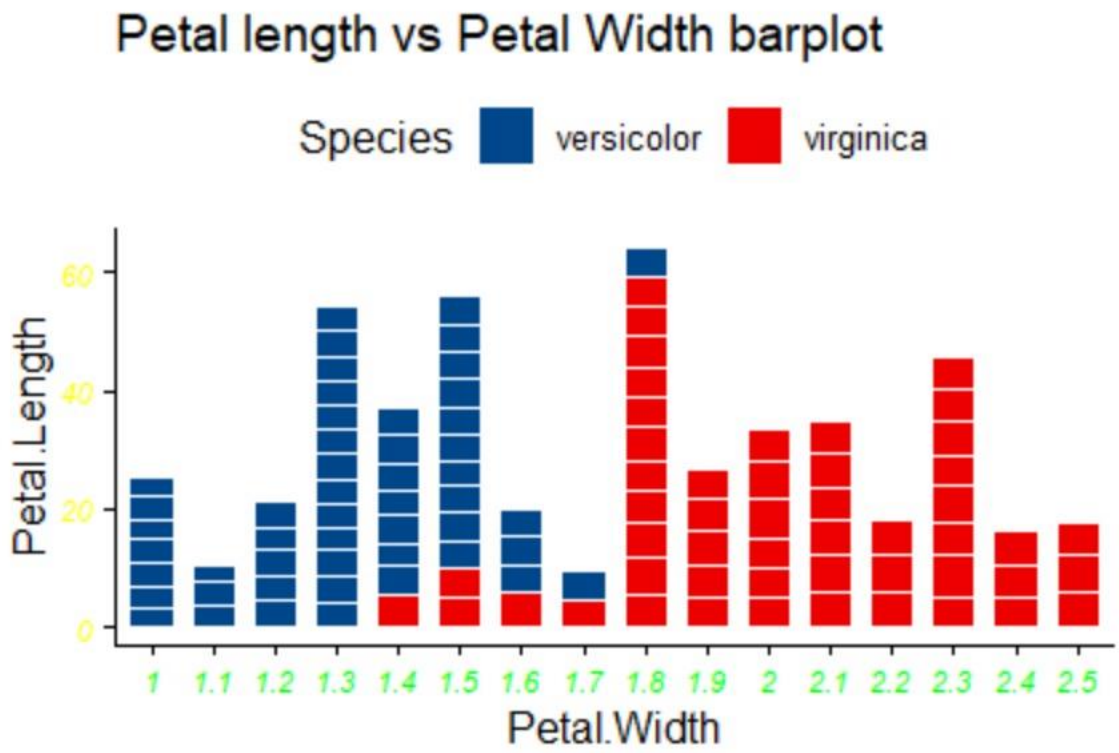
$$b_j(x_i)= I(c_j \le x_i \ c_{j+1}) $$

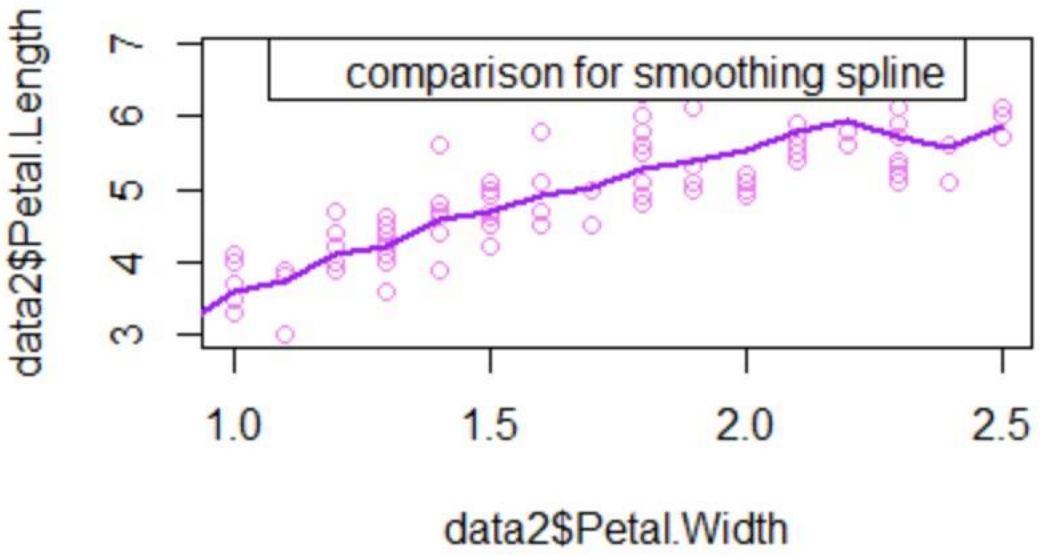
```{r setup, include=FALSE}
library(splines)
knitr::opts_chunk$set(echo = TRUE)
data(iris)
data1 =data.frame(iris)
data1
data2<-data1[!(data1$Species=="setosa"),]
data2
fit1<-smooth.spline(iris$Petal.Width,iris$Petal.Length,df=16)

fit1
fit2<-smooth.spline(iris$Petal.Width,iris$Petal.Length,cv=TRUE)
fit2


library("ggpubr")
head(data2[, c("Sepal.Length", "Sepal.Width", "Petal.Length","Petal.Width", "Species")])

plot(x=data2$Petal.Width, y=data2$Petal.Length, col="violet")

lines(fit1,lwd=2,col="purple", main = "comparison for smoothing spline")
legend (x="top",legend=c("comparison for smoothing spline"),
col=c("red","blue"))
```

```
s <- ggbarplot(data2, x = "Petal.Width", y = "Petal.Length",
        fill = "Species",
        color = "white",
        palette = "lancet",
        main = "Petal length vs Petal Width barplot",
        fig.cap("Hello"))
s
s+theme(axis.text.x = element_text(color = "green",size = 8,face = "italic"),axis.text.y =
element_text(color = "yellow",size =8,face = "italic"))


legend (x="topleft",legend=c("Petal length vs Petal Width comparison for smoothing
spline"),
col=c("red","blue"))
```

2) a)The optimal value of lamda for smoothing spline
is 0.000781036

b) The above code gives the publication ready figures and curves are plotted below






```{r setup, include=FALSE}
library(gam)
knitr::opts_chunk$set(echo = TRUE)
set.seed(1)
require(ISLR);
library(leaps)

data(College)
training <- sample(1: nrow(College),0.8* nrow(College))
trainingData <-College[training,]
testing <-College[-training,]
model <- regsubsets(Outstate ~., trainingData ,method='forward',nvmax = 17)


fit <-  gam(Outstate ~ Private + s(Room.Board, df=5) + s(PhD, df=5) + s(perc.alumni, df=5) +
            s(Expend, df=5) + s(Grad.Rate, df=5),data=trainingData)



```

3) a)From the above code we splitted into training and test data set and fitted GAM for training dataset.

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
par(mfrow=c(2,3))
```

```
plot(fit, se=TRUE, col="red")

summary(fit)
```
b)Here we plot GAM and From the non parametric Anova analysis we can conclude that expend shows the strong relationship between response and expend,moderate relationship between response and phd or reponse and Grad.Rate

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
testset <- predict(fit,testing)
error1 <- mean((testing$Outstate -testset)^2)
error1

tss1 <-mean((testing$Outstate-mean(testing$Outstate))^2)

R_test <- 1- error1/tss1
R_test


trainset <- predict(fit,trainingData)
error2 <- mean((trainingData$Outstate -trainset)^2)
error2

tss2 <-mean((trainingData$Outstate-mean(trainingData$Outstate))^2)

R_train <- 1- error2/tss2
R_train

overfitting <- R_test - R_train

overfitting


```

c) from the above code we got the results for R square for testing is 0.78 and R square for training is 0.79 .Here ,training and testing error are high ,so it is not a overfitting model.

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```{r cars}
summary(cars)
```

Including Plots

You can also embed plots, for example:

```{r pressure, echo=FALSE}
plot(pressure)
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.